

Klamath Network Data Mining Phase II Protocols



**Laura Bridy, Elizabeth Perry, Tim Shepherd and Robert Truitt
Klamath Network
1250 Siskiyou Blvd.
Ashland, OR., 97520-5011**

November 23, 2005

Table of Contents

Klamath Network Data Mining Phase II Protocols.....	1
Table of Contents	2
Table of Figures	2
1.0 Introduction.....	3
2.0 Objectives	4
3.0 Protocols	4
3.1 Cataloging Datasets	4
3.2 Cataloging References	7
3.3 Documenting Additional Taxa	8
4.0 Priorities.....	8
4.1 Spatial Datasets Protocol	8
5.0 Where Does the Information Go After it is Created?.....	9
6.0 Conclusions and Recommendations.....	10
7.0 References.....	10
7.1 Additional Useful References	10
Appendix A – The Metadata Interview	12
Appendix B – Protocol for Naming Metadata Files.....	12

Table of Figures

Figure 1: Dataset Catalog and Metadata Work Flow.....	5
---	---

1.0 Introduction

With the issuance of the Natural Resource Challenge (NRC) by Robert G. Stanton (1999), the NPS Inventory and Monitoring Program (I&M) responded by creating databases designed to catalog the natural resource knowledge in each park unit. These databases (NatureBib, NPSpecies, NPS Data Store, and Dataset Catalog) will give park managers, incoming park staff, and researchers' access to the scientific information needed to effectively manage the park.

In our data mining efforts, we are looking to catalog the natural resource information held at each park. The cataloging of research and management projects conducted at each park unit in the Klamath Network will help preserve the institutional memory and will allow future studies to build upon past projects.

The natural resource information is classified into 12 basic inventories, according to Inventory and Monitoring program guidelines ([NPS, 2003](#)):

- 1) Species occurrence (comprehensive list of vertebrates and vascular plants for each park).
- 2) Species abundance and distribution (vertebrates and vascular plants).
- 3) Natural Resource Bibliographies.
- 4) Water Quality (EPA's STORET DB).
- 5) Geology.
- 6) Soils.
- 7) Air Quality.
- 8) Air Visibility.
- 9) Location of Water Bodies.
- 10) Climate (Weather).
- 11) Base Cartography.
- 12) Vegetation Mapping.
- 13) Sound (unofficial 13th inventory).

During the first phase of data mining in the Klamath Network, we concentrated our efforts on six taxonomic groups: amphibians, birds, fish, mammals, reptiles, and vascular plants. Within these groups, we found documentation to support a species list, located documents with species abundance and distribution, and used the documents to populate the NatureBib bibliography and NPSpecies databases.

In Phase II, we will expand our criteria to include documents from the other inventory categories. These documents will be cataloged as in Phase I, with the addition of cataloging all taxonomic references as well. Another aspect of Phase II is to document datasets¹ using the I&M Access database, Dataset Catalog. This aspect requires basic information about the data (purpose of dataset, data collection dates, methods of data collection, etc.). If enough information about the dataset is available, Federal Geographic Data Committee (FGDC)

¹ We recognize that "dataset" is also commonly written as "data set". Although the two spellings are interchangeable, we have decided to use "dataset" in order to maintain consistency.

compliant metadata² will be created in NPS Metadata Tools. Data with spatial components that may be utilized for KLMN Inventory and Monitoring projects that are not already in GIS will be converted into GIS, if enough spatial information is available. Lastly, the data may be incorporated into the Klamath Network Natural Resources Database (KLMN_NRD).

2.0 Objectives

As a continuation of the steps performed in Phase I of the Klamath Network's data mining efforts, Phase II will attempt to more fully capture and catalog the natural resource data in the six network parks. To do so, the cataloging criteria must be expanded. This expansion of criteria results in the following objectives:

- 1) Cataloging datasets relating to all of the 12 basic inventories,
- 2) Cataloging references relating to the remaining 12 basic inventories not addressed in Phase I, and
- 3) Cataloging references and datasets containing taxonomic information beyond vascular plants and vertebrates.

The resulting goal of finding, capturing, and cataloging the information (natural resource documents, maps, and datasets) in Phase II is to have these resources accessible to managers, aiding in the decision-making process. After we complete as much of Phase II in each park as time permits, the information cataloged will be sent to the Network Data Manager and to each of the contributing researchers in the parks. The public will eventually also be able to access the information cataloged in the online databases.

The emphasis will be on data that is viewed as a priority for the parks and completing basic cataloging of this information.

3.0 Protocols

3.1 Cataloging Datasets

In order to accomplish the first objective of cataloging datasets, we will utilize Dataset Catalog and NPS Metadata Tools. Before using these programs, we will approach each park's Resource Management staff and ask each branch to prioritize their data for entry into the Dataset Catalog database (see Appendix A: Metadata Interview). This gives us the opportunity to ask permission to access the appropriate network servers and personal computer hard drives and also to ask where the data files and documents connected to the data are located. One caveat is that digital data already in GIS applications (e.g., ArcView 3.x or ArcGIS 9.x) will be noted but not considered at this time. After locating the files, we can start the cataloging and metadata process (see Figure 1).

² “‘Metadata’ is information about data or other information. They describe the ‘who, what, where, when, why, and how’ of every aspect of the data.” (USGS, 2005)

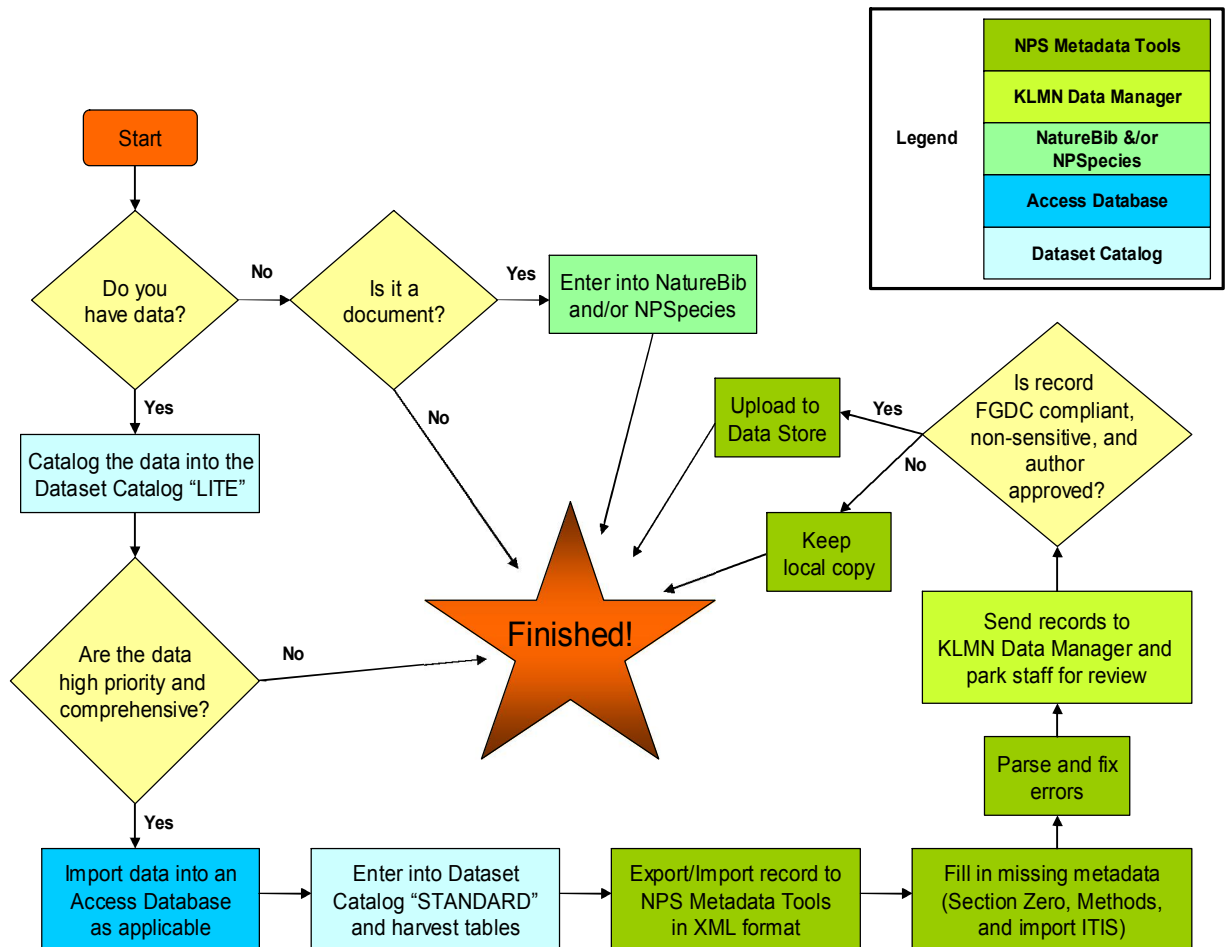


Figure 1: Dataset Catalog and Metadata Work Flow

3.1.1 Dataset Identification

Once data are located and identified, we will create a record in Dataset Catalog for each dataset that was part of the study. Steve Tessler and Joe Gregson (1997) of the NPS I&M Program provide an elegant definition of a dataset in their “[Draft Data Management Protocol](#)” website ([Chapter 3, Section Data set Identification](#)), which is excerpted below.

The following three statements and examples help define what constitutes a separate line (an eventual data set [sic] catalog entry) in the data resource list. The criteria also suggest when to combine entities that share the important traits of objective, methodology, or investigator.

(1) A single data set [sic] in the catalog contains data collected for a single objective.

Example: The protocol for monitoring aquatic macroinvertebrates calls for taking insect samples, measuring discharge, recording water chemistry parameters, and performing a suite of habitat measurements. Whereas the meat of the project is getting the bugs, the other data are important companions and may be critical to the interpretation of patterns

of diversity and community in the insect data. However, the companion data are not unique to macroinvertebrate monitoring, so the bug data stands alone.

The chemistry, discharge, and habitat data also stand alone. Depending on the objectives of the field foray, one, two, or all could be sought with or without biological data. Stream chemistry measurements are essentially the same thing (and viewed the same) whether the task is collecting bugs, shocking ponds, or demonstrating techniques during I&E [sic] training sessions. It is a separate data/information entity.

(2) Where the same information objective is pursued using different methodologies, each should be cataloged separately but share Subject and Keyword elements.

Example: Air quality sampling in Shenandoah National Park provides an example of measuring visibility in the park with camera, transmissometer, and the IMPROVE (multi-agency) devices and protocols. Each data set [sic] tells something about visibility, but their methods, time intervals, and resulting data sets [sic] are distinct; thus, each gets its own catalog entry. Using the same keyword(s) (e.g., visibility) identifies *all three* data sets and allow the catalog explorer distinguish usability for particular needs.

A second example is the three bird census methodologies in Shenandoah National Park: the MAPS program (net capture), Breeding Bird Survey (point counts), and transect surveys (visual and voice). Although each bird census results in complementary but different kinds and qualities of data; each is cataloged separately.

(3). If the same kind of data are gathered by separate investigations or different SOPs, each data set [sic] gets its own catalog entry.

Example: Stream chemistry parameters are measured by park staff during resource monitoring, wastewater effluent evaluation, and potability studies. Even if all sought only temperature, pH, and DO (dissolved oxygen), the difference in who collects the data is significant and can reflect real differences in equipment, methodology, training, accuracy, and resolution. Each gets its own catalog entry. This redundancy is instructive and may indicate the need to consolidate these activities or at least standardize the equipment and procedures.

In another example from Shenandoah National Park, water chemistry characterization is the target of several studies by the University of Virginia. *Whenever* an outside investigator generates the data it is a separate catalog entry. Again, all of these studies share the bulk of their keyword terms and is identified as a group during any query of water chemistry studies in the park.

Each related dataset will be connected through the "related data" tab in the Dataset Catalog database. Other ways they can be related, as necessary, include supporting documents (publications, annual reports, etc. via the NatureBib BibKey_ID), keywords, and title. Information on and the Dataset Catalog database application are available at the [I&M Dataset Catalog Web Page](#) (2005).

Dataset Catalog has two "views" for entering information about the data. The "lite" view is what NPS considers minimum metadata, and the "standard" view is more comprehensive, which

approaches compliant metadata standards (FGDC sections 1 and 7). The “view” that is used for cataloging the data will depend on a couple of factors: 1) Is the data a priority for the parks, 2) how comprehensive are the data, 2) how comprehensive is the available information about the data, and 3) how feasible is it to create metadata, given the amount of data and time permitted at that park. It should be noted that there are varying levels of natural resource information in the six network parks.

The emphasis will be on using the “lite” view to capture as many datasets as possible. For the higher priority datasets, the goal is to use the “standard” view in Dataset Catalog. In order to create complete metadata, we will import data into an Access database if necessary and feasible. This conversion allows the Dataset Catalog to “harvest” the data’s table information (column headings, attributes, validation rules, etc.). The harvested table information is required for compliant metadata.

When the “standard” view is complete in Dataset Catalog, we will export each record as an .xml file (the standard format for metadata files). We will then import this file into NPS Metadata Tools to capture information not contained in Dataset Catalog.

NPS Metadata Tools can document the taxonomic classification of species in a dataset. We import the species’ taxonomic classification(s) from the [ITIS](http://www.itis.gov/) website (<http://www.itis.gov/>, 2004) into NPS Metadata Tools. In the event that the scientific name is not available in ITIS, we will first attempt contacting the investigator to verify that the name is correct, to try to get an ITIS approved scientific name, to obtain the reason why they chose a different name, etc. If the investigator is not available, there are two options:

- 1) Change the import list to ITIS if we are absolutely positive of the identification, spelling, etc. that the investigator meant. Document every change that we made in the metadata and that the investigator could not be contacted.
- 2) If we are the slightest bit uncomfortable with changing the list to import to ITIS, we will not fix the questionable name(s) and we will stop. This will happen in the majority of cases. Document in the metadata that the biological hierarchy is not complete because of non-matches in ITIS and that the investigator could not be contacted.

Scientific names that do not match ITIS currently cannot be imported, even if these are the preferred names by the investigator. The investigators may be able to provide their preferred hierarchies. Once again, we will document these names in the metadata so the information is not lost. We will add the non-matched names into the “Comments” field of Dataset Catalog, or the “Supplemental Information” field of NPS Metadata Tools. The metadata will later be checked for errors (parsing) and corrected if needed.

3.2 Cataloging References

In order to accomplish the second objective of cataloging references, we will enter documents into the NatureBib and NPSSpecies applications as was done in Phase I (Smith and Truitt, 2005). During the Phase I process, the focus was on water quality, vascular plants, and vertebrate species information. Phase II will shift focus predominately in the discovery and cataloging of references concerning the remaining 12 inventories (e.g., air, climate, geology, and soils).

Concerning digital documents, NatureBib holdings locations will be recorded as: “DIGITAL LOCATION: S:\team\...” etc. NatureBib numbers will be recorded on the actual documents with park permission:

- For Microsoft Word documents, the NatureBib numbers will be recorded in the “File,” “Properties,” “Keywords” section.
- For files in PDF format, the NatureBib numbers will be recorded in the “File,” “Document Properties,” “Keywords” section with Adobe Acrobat Professional.

By inserting the NatureBib number in this manner, we are able to identify which documents have been entered, even if their file locations change. This method of inserting the information also ensures that the NatureBib numbers are searchable and the printed document will only contain our notation if the viewer chooses. NatureBib numbers will be recorded on hardcopy documents with the same method as Phase I.

3.3 Documenting Additional Taxa

In order to accomplish the third objective of documenting additional taxa, we will capture invertebrate and non-vascular plant taxonomic references and datasets as we come across them using the same process as we did in Phase I. This is not a priority of Phase II but will be accomplished when we discover references associated with this information.

4.0 Priorities

Priorities will be based upon a number of pre-activities that the Data Mining Team (DMT) will perform at each park to determine the breadth, extent, and priority of datasets available. First, the DMT will work with their park Point of Contact (POC) and/or the Natural Resource Chief to obtain a general idea of the number of digital and hardcopy datasets that are available within a park as well as the scope of those datasets. The scope will be datasets that are predominately species-related (i.e., wildlife and botanical), physical sciences, or some combination of both. Second, the DMT will query the park’s Natural Resource staff as to where the majority of the datasets reside (e.g., personal computer hard drives, network servers, optical discs, etc.) and distribute a metadata questionnaire (see Appendix A). Lastly, the DMT will collate the information obtained concerning the datasets and, with the park’s POC or Natural Resource Chief, prioritize the list of datasets.

In each branch, the focus will be on producing at least a basic level of metadata for as many datasets as possible; creating full metadata on complete datasets is secondary. The quantity of metadata created will depend on the quality of each dataset and time constraints.

4.1 Spatial Datasets Protocol

If we find data that have a spatial component but are not in GIS applications, we will send them to the Klamath Network GIS Specialist for conversion into ArcCatalog and upload to the NPS Data Store, as appropriate. The criteria for sending datasets to be uploaded into NPS Data Store are that the dataset: 1) could contribute to KLMN “Vital Signs” Monitoring, 2) has a spatial component that can be referenced to a specific coordinate system and 3) has a datum³ or projection, which links it to a specific place on earth. Two examples of acceptable criteria are

³ A datum is the reference point on earth to which the coordinate system is based. Typical examples of datum are NAD27 or NAD83, referring to the North American Datum of 1927 or 1983.

location information in the Geographic System (latitude and longitude coordinates) or Universal Transverse Mercator (UTM) coordinates, both of which must have a datum. Without a datum, there can be at least 200 meters of error.

If a dataset without a datum contains many location points and is deemed valuable by park management, then we will send it to the Klamath Network GIS Specialist. The level of error will be documented in the metadata. Datasets with very few spatial components (e.g., four or five points) and no datum will be sent only if they are deemed critical by park management. An example is a few location points (without datum information) concerning an endangered species.

Location points can be in a variety of coordinate systems, including latitude and longitude; UTM; decimal degrees; and the Public Land Survey System of township, section, and range.

When the datasets are sent, they should be in the format of Excel, Access, text, or another file type that can be easily imported into these formats. These spatial datasets should be emailed to the Klamath Network GIS Specialist in groups with each dataset in a separate folder. Each dataset folder should contain the complete dataset, its non-spatial metadata, and a brief description of the spatial data and its location.

For more details about the spatial references protocol, contact the Klamath Network GIS Specialist. For more details about spatial references in general, see the [Consortium for International Earth Science Information Network](#) website (1998).

5.0 Where Does the Information Go After it is Created?

NatureBib and NPSpecies entries are created online and available as soon as they are created to park and network personnel via the internet. This service will also eventually be available for people outside the NPS, with appropriate limitations.

The metadata have a longer, more complicated process. This process involves steps at the park, network, and national levels.

At the park level: all records will be saved on a network drive. For each investigator whose data we catalog, we will also provide a digital copy of the finished product. This may be a CD, DVD, portable hard-drive, or network copy, depending on staff preference and the amount of data. We will include metadata and spreadsheets describing data found (including data too obscure to be cataloged, GIS notes, species notes, etc.). Please refer to Appendix B for the protocol on naming the metadata files we create.

At the network and national levels: in the beginning of Phase II, we will forward completed Dataset Catalog and NPS Metadata Tools files to the Klamath Network Data Manager via email. After verifying our competence with the metadata creation process, CDs of our work will be sent to the Data Manager monthly. The Data Manager will determine whether the metadata are uploaded to the NPS Data Store or kept locally. This will depend on the completeness of the metadata (only FGDC compliant metadata can be uploaded), the sensitivity of the data, and the investigator's preference. Data miners will upload appropriate metadata. Raw data will not be uploaded, except for base cartography.

6.0 Conclusions and Recommendations

In completing Phase II of data mining, we aim to have each park's natural resource inventories documented in the appropriate database: datasets recorded in Dataset Catalog (eventually linked to NatureBib), documents cataloged in NatureBib, and species lists compiled in NPSpecies (linked to NatureBib). Given enough information on the data (and time at that park), Dataset Catalog and NPS Metadata Tools will be used to create FGDC compliant metadata. Ideally, a mechanism would be in place at each park to capture basic information about each new project or study to be done within the park. It is envisioned that all natural resource data will be placed into a standardized database (such as the KLMN_NRD), cataloged, and easily accessible to researchers and park managers.

7.0 References

- Consortium for International Earth Science Information Network. (1998, Mar.). Spatial Reference. <http://www.ciesin.org/metadata/documentation/guidelines/spatref.html>. (Nov. 23, 2005).
- National Park Service. (8 Oct. 2003). NPS: Nature & Science: Inventory and Monitoring of Park Natural Resources. <http://www.nature.nps.gov/protectingrestoring/IM/resourceinventories.cfm>. (Dec. 7, 2005).
- Inventory & Monitoring Program. (2005, Nov. 9). Dataset Catalog Desktop Applications. <http://science.nature.nps.gov/im/apps/datacat/index.htm>. (Nov. 18, 2005).
- ITIS. (2004, Nov. 24). Integrated Taxonomic Information System. <http://www.itis.gov/>. (Nov. 20, 2006).
- Smith, S. and Truitt, B. (2005). Klamath Network Data Mining Protocol, KBaM!.
- Stanton, Robert G. (1999, Aug. 12). Natural Resource Challenge. "Re-birth of a Park": Mount Rainier National Park Centennial. http://www.nature.nps.gov/challenge/challengedoc/index_frames2.htm. (Nov. 14, 2005).
- Tessler, S. and Gregson, J. (1997). Chapter 3 – I&M Data Set Catalog. Draft Data Management Protocol. <http://www1.nrintra.nps.gov/im/dmproto/joe40001.htm>. (Nov. 1, 2005).
- USGS. (2005, Nov. 4). USGS CMG "Formal Metadata" Definition. Coastal & marine Geology InfoBank. <http://walrus.wr.usgs.gov/infobank/programs/html/definition/fmeta.html>. (Nov. 22, 2005).

7.1 Additional Useful References

- Dataset Catalog: <http://science.nature.nps.gov/im/apps/datacat>.
- Federal Geographic Data Committee (FGDC): <http://fgdc.er.usgs.gov/fgdc/fgdc.html>.
- Inventory and Monitoring Statement of Purpose, NPS-75: www.nature.nps.gov/nps75/nps75.pdf
- Metadata Authoring Guidance: <http://science.nature.nps.gov/nrdata/docs/metahelp/NR-GISMetadataAuthoringGuidance.pdf>.
- NPS Inventory and Monitoring Home Page: <http://science.nature.nps.gov/im/index.htm>.
- NPS Metadata Tools and Editor: <http://science.nature.nps.gov/nrgis/tools/editor.cfm>.

Spatial Reference Descriptions:

<http://www.ciesin.org/metadata/documentation/guidelines/spatref.html>.

USGS Formal metadata: information and software: <http://geology.usgs.gov/tools/metadata/>.

Appendix A – The Metadata Interview

Metadata Project

by the Klamath Network

“As part of the Service’s efforts to ‘improve park management through greater reliance on scientific knowledge,’ a primary purpose of the Inventory and Monitoring Program is to develop, organize, and make available natural resource data and to contribute to the Service’s institutional knowledge by facilitating the transformation of data into information through analysis, synthesis, and modeling” (I&M Program, NPS 2004).

To achieve this goal, we (the Klamath Network Data Mining Team) will catalog natural resource data collected at each of the six national park units associated with the network (Crater Lake NP, Lassen Volcanic NP, Lava Beds NM, Oregon Caves NM, Redwood NSP, and Whiskeytown NRA). We are placing this information into an Access database and creating a catalog of key descriptive information about the data. Without the catalog, crucial facts about the data may be lost. We are including all datasets because each one may have current or future relevance to the I&M Vital Signs Monitoring Program. The cataloged information will be made available both to the park personnel and the public. Each park will be supplied with a full copy of the completed Dataset Catalog. The public will have limited access through the uploading of the Dataset Catalog to the NR-GIS Data Store by the network’s data manger. The level of availability will depend on the completeness of the metadata, your preference, and the sensitivity of the information (e.g., T&E species). Also, the data or dataset will not be posted at the NR-GIS Data Store; only information “about” the data and who to contact if further information about the dataset is desired. The decision whether or not to make the dataset available will rest with the park and researcher(s).

We need your help identifying the most important data to catalog. Also, we would appreciate examining associated summaries and reports (progress reports, published reports, etc.) that go with each set of data. Please let us know the locations of these documents.

To capture essential descriptions of crucial elements of the data, please fill out the questionnaire on the following pages.

Metadata Interview Questionnaire

Please direct us to the most important datasets first.* If the answers to some of the questions can be answered with the dataset itself, or existing supporting documentation such as summaries and reports, please direct us to that information and skip to the next question.

Your name _____

1. Where is the dataset (e.g., Network Drive S:\Team\Veg\Sillett_tree_physiology.mdb)?
2. What is the **location** of the study? What areas of Redwood National and State Parks? Please include the names of any other **National Parks** in which the study was conducted.
3. What is the **title** of the dataset?
4. Who is the **author** of the dataset?
5. What is the **date** of the latest version (published or made available for release)?
6. What is the **permit number** (e.g., NPS Research Permit and Reporting System)?
7. What is the **project name**?
8. What is the **project number** (e.g., PMIS or RMP project number)?
9. Are there any **keywords** that you especially want to describe the dataset (e.g., northern spotted owl, *Strix occidentalis caurina*, abundance, Bald Hills)?
10. Please give a brief **abstract** of the dataset. Include information such as the project from which the data are derived, who was involved, general methodology used (#sites, sampling frequency, protocol, equipment), and references to concurrent or related data. If this information is available in a supporting document, please direct us to it, list the location below, and skip the rest of this question. If writing an abstract here, use extra pages if necessary.

11. Briefly, what is the **purpose** of the dataset (i.e., why was the data collected, and what use(s) or information will the data provide)?

12. Please indicate the **timeframe** of data collection in the most exact date possible:

☐ Single date: on _____
☐ Span of dates: from _____ to _____
☐ Multiple dates: (please list) _____

13. How **frequently** is the dataset **updated** (i.e., the interval at which new data are appended to the dataset)? Pick one:

☐ Continually ☐ Daily ☐ Weekly ☐ Monthly ☐ Annually ☐ Biannually ☐ As needed
☐ Irregular ☐ None planned ☐ Unknown

14. What is the **status** of the dataset? Pick one:

<input type="checkbox"/> New	a dataset in the planning, implementation, or collection stage
<input type="checkbox"/> Active	data are still being added to the dataset periodically
<input type="checkbox"/> Inactive	data is no longer being collected but may have future updates
<input type="checkbox"/> Legacy	the data were collected by previous projects or personnel that needs validation and documentation
<input type="checkbox"/> Partial	a dataset that is in work or that has not been completed per data gathering protocols/specifications
<input type="checkbox"/> Historic	data without planned updates from historical natural resource activities
<input type="checkbox"/> Other	

15. What **progress** has been made on this dataset?

☐ Planned ☐ In work ☐ Complete

17. Please briefly explain the **quality** of the data. Specifically, please address any issues that would affect data quality; whether or not the data has been **verified, validated**, and/or critically reviewed; and if the data were created according to a set standard.

18. Does this dataset contain **sensitive** information? ☐ Yes ☐ No

If **yes**, please specify who should be able to view this metadata:

☐ Public access denied ☐ Federal only ☐ NPS only ☐ Park only

19. Please describe the **data columns/fields**. Give meanings for abbreviations (ex: dbh = diameter at breast height, gen = genus, Mamu = marbled murrelet, *Brachyramphus marmoratus*). Please give scientific names for species. If this information is located in the data or elsewhere, such as a "data dictionary," please direct us to it, and skip the rest of this question.

20. Do you think this **descriptive information about the data** should be uploaded to NR-GIS DataStore, where it will be accessible to the general public? ☐ Yes ☐ No

21. Should the **data** be uploaded to NR-GIS DataStore, where it will be publicly accessible?
☐ Yes ☐ No

22. Please briefly list any **additional** important information about the dataset that has not been captured elsewhere on this form.

*"The most important datasets" are those that are the most important to share with other researchers, and the most important to carry on to the next generation in perpetuity.

Citations

I&M Program, National Park Service. (August 3, 2004, Draft). I & M Data Management Vision and Framework. Inventory & Monitoring Program.
<http://science.nature.nps.gov/im/datamgmt.htm>. (November 1, 2005).

Appendix B – Protocol for Naming Metadata Files

The automatically generated metadata .xml and .txt file names are not too informative as to the content of the files, nor are they standardized in a manner consistent among copies of the Dataset Catalog and NPS Metadata Tools databases. To make the names more informative as well as standardize the naming structure, the Klamath Network has decided to consistently change the metadata file names immediately after exporting.

File names will follow the structure:

PARKCODE_Branch_MajorClass_ShortTitle_Author_Year.xml.

PARKCODE = the four letter NPS park unit code.

Branch = the division within the park from which the data originates.

Major Class = the four letter code of the major category under which the dataset falls. The four letter codes (from the KLMN-NRD) we will use are:

Four Letter Code	Major Category
AIRQ	Air Quality
BIRD	Birds
CLIM	Climate/Weather
FIRE	Fire
FISH	Fish
GEOL	Geology
H2OQ	Water Quality
HERP	Herpetology
INVT	Invertebrate
MAMM	Mammals
NPLT	Non-Vascular Plants
SOIL	Soils
SOUN	Sound
VPLT	Vascular Plants

ShortTitle = a brief descriptive name of the dataset (i.e., an abbreviated descriptive form of the actual dataset title).

Author = the last name of the dataset's author if it is not elsewhere in the file name and is an actual person or organization (not the branch, park, etc.).

Year = the publication date or for long-term monitoring, the year the metadata file is created, if this information is not conveyed elsewhere in the file name.

Notes:

No spaces or periods will be used in the file name; only underscores will be used.

Different words within ShortTitle will be differentiated by capital letters starting each new word.

When the author's name and the year are unnecessary, the file name will read:

CODE_Branch_MajorClass_ShortTitle.xml.

When the author's name is unnecessary but the year is important, the file name will read:

CODE_Branch_MajorClass_ShortTitle_Year.xml.

When the year is unnecessary but the author is important, the file name will read:

CODE_Branch_MajorClass_ShortTitle_Author.xml.